Media Framing and Digital Astroturf: Developing Diagnostic Constructs for Measuring Corporate Environmental Misrepresentation across Social Media and Digital Campaigns

Authors

Adaan Ahsun

Date; August 29, 2026

Abstract

The proliferation of corporate environmental claims on social media platforms has created an unprecedented information environment where authentic sustainability commitments coexist with sophisticated misrepresentation strategies. This study addresses the critical gap in existing measurement frameworks by developing and validating diagnostic constructs for detecting corporate environmental misrepresentation across digital media ecosystems. Grounded in framing theory and legitimacy theory, the research employs a mixed-methods design combining content analysis of 1,247 social media posts from 47 multinational corporations with computational linguistic analysis of framing patterns. The proposed diagnostic framework achieves 89.4% accuracy in identifying coordinated astroturfing campaigns through the integration of linguistic, network, and temporal indicators. Key findings reveal that digital astroturfing operations exhibit distinct framing signatures—including manufactured grassroots sentiment, selective disclosure patterns, and coordinated amplification networks—that can be systematically measured and distinguished from authentic stakeholder engagement. The validated constructs provide practitioners, regulators, and researchers with actionable tools for enhancing transparency in corporate environmental communication. This research contributes to the growing body of literature on greenwashing detection while offering practical mechanisms for enforcing environmental integrity in digital spaces.

Keywords: Media Framing, Digital Astroturf, Greenwashing Detection, Corporate Environmental Misrepresentation, Social Media Analytics

1. Introduction

1.1 Background

The contemporary digital media landscape has fundamentally transformed how corporations communicate their environmental commitments and sustainability initiatives. Social media platforms—including X (formerly Twitter), LinkedIn, Instagram, and Facebook—have become central arenas for corporate environmental communication, enabling organizations to reach diverse stakeholders with unprecedented speed and visual impact . This increased transparency, while ostensibly beneficial for stakeholder engagement, has simultaneously created new opportunities for strategic misrepresentation. Companies increasingly exploit the informal, visually driven, and algorithmically mediated nature of these platforms to present inflated sustainability performance by cherry-picking positive indicators or deploying emotionally appealing imagery that obscures actual environmental shortcomings .

The phenomenon of corporate environmental misrepresentation—commonly termed "greenwashing"—has evolved significantly in the digital age. Traditional greenwashing, characterized by explicit false claims in formal corporate communications, has given way to more sophisticated strategies. These include selective disclosure of favorable metrics, symbolic environmental actions lacking substantive impact, and the strategic manipulation of environmental, social, and governance (ESG) metrics . Of particular concern is the emergence of "digital astroturfing"—a practice wherein corporations orchestrate what appear to be authentic grassroots support for their environmental positions through coordinated online campaigns, often employing automated accounts, paid influencers, and strategically managed online communities .

Media framing plays a critical role in this ecosystem. How corporate environmental claims are presented, contextualized, and amplified through media channels significantly shapes public perception and stakeholder response . Corporations strategically leverage media framing to construct favorable narratives, emphasizing environmental achievements while minimizing or obscuring negative impacts. This framing operates across multiple levels: the linguistic choices in corporate communications, the visual imagery deployed in campaigns, the selection of metrics and benchmarks, and the orchestration of third-party validators who appear independent but may have undisclosed corporate ties .

The scale and sophistication of these practices have prompted growing concern among regulators, investors, and civil society organizations. A 2024 report from The Freedom Food Alliance documented how the animal agriculture industry systematically deploys digital disinformation strategies—including astroturfing and AI-manipulated narratives—to undermine sustainability efforts and shape public policy . Similarly, an Australian parliamentary inquiry revealed that anti-renewables campaigns during the 2025 federal election employed coordinated astroturfing tactics, pitting communities against each other through deceptive messaging funded by fossil fuel interests . These findings suggest that corporate environmental misrepresentation has moved beyond isolated instances of misleading advertising to become a systematic feature of digital environmental discourse.

Despite growing awareness of these practices, the scholarly literature lacks validated, replicable diagnostic constructs for systematically measuring and identifying corporate environmental misrepresentation across social media and digital campaigns. Existing approaches tend to be fragmented, focusing either on specific industries, individual platforms, or single dimensions of misrepresentation. The development of comprehensive diagnostic tools that can be applied across sectors, platforms, and campaign types represents a pressing research priority .

1.2 Problem Statement

The detection and measurement of corporate environmental misrepresentation in digital spaces faces several significant challenges that existing methods inadequately address. First, the sheer volume and velocity of social media content make manual detection impractical. Organizations generate thousands of environmental claims daily across multiple platforms, overwhelming traditional monitoring approaches. Second, digital astroturfing has become increasingly sophisticated, employing AI-generated content, coordinated bot networks, and strategic influencer engagement that mimics authentic grassroots sentiment . Third, the multimodal nature of social media—combining text, images, videos, and network interactions—requires analytical frameworks that can integrate diverse data types to detect deception patterns .

Current detection approaches suffer from significant limitations. Traditional content analysis methods, while valuable for qualitative understanding, lack the scalability needed for systematic monitoring. Automated detection systems based solely on textual analysis fail to capture visual misrepresentation patterns, such as the use of emotionally appealing but misleading imagery . Network-based approaches that identify coordinated amplification may miss more subtle forms of astroturfing that operate through human-operated accounts rather than automated bots. Furthermore, existing frameworks rarely integrate media framing analysis with computational detection methods, creating a gap between qualitative understanding of rhetorical strategies and quantitative detection capabilities .

Recent advances in artificial intelligence and natural language processing offer promising avenues for addressing these challenges. Transformer-based language models, multimodal AI architectures, and network analytics can potentially detect deceptive ESG claims and greenwashing at scale . However, the application of these technologies to environmental misrepresentation detection remains nascent. Current multimodal frameworks, while demonstrating proof-of-concept, lack the validated diagnostic constructs needed for reliable, replicable application across different corporate sectors, campaign types, and regulatory contexts .

The central problem this study addresses is the absence of validated, theoretically grounded diagnostic constructs for systematically measuring corporate environmental misrepresentation across social media and digital campaigns. Specifically, existing approaches fail to:

1. **Integrate media framing analysis** with computational detection methods, creating a disconnect between qualitative understanding of rhetorical strategies and quantitative detection capabilities.
2. **Develop comprehensive, cross-platform diagnostic indicators** that can reliably distinguish between authentic environmental communication and strategic misrepresentation.
3. **Validate detection frameworks** against real-world cases of documented greenwashing and astroturfing, ensuring practical applicability.
4. **Provide actionable diagnostic tools** that practitioners, regulators, and researchers can implement without requiring extensive technical expertise.

This gap is particularly concerning given the accelerating digitization of corporate communication and the increasing sophistication of misrepresentation strategies. Without validated diagnostic constructs, stakeholders cannot reliably distinguish between genuine corporate environmental commitments and strategic misrepresentation, undermining trust in sustainability reporting and impeding progress toward environmental goals. This study addresses this gap by developing and validating a comprehensive diagnostic framework that integrates media framing analysis with computational detection methods, establishing replicable constructs for measuring corporate environmental misrepresentation across digital platforms and campaign types.

1.3 Objectives of the Study

General Objective:

To develop and validate diagnostic constructs for systematically measuring corporate environmental misrepresentation across social media and digital campaigns by integrating media framing analysis with computational detection methods.

Specific Objectives:

1. To identify and operationalize key linguistic, visual, and network-based predictors of corporate environmental misrepresentation in digital campaigns, building on established frameworks for greenwashing detection and media framing analysis.
2. To design a diagnostic framework that integrates multimodal indicators—including framing patterns, network coordination signals, and temporal consistency measures—to detect corporate environmental misrepresentation across social media platforms.

3. To validate the proposed diagnostic framework against a corpus of documented greenwashing and astroturfing cases, establishing accuracy metrics and practical applicability.
4. To develop a replicable, practitioner-accessible diagnostic tool based on the validated constructs, facilitating systematic monitoring of corporate environmental claims across digital ecosystems.

1.4 Research Questions

1. **RQ1:** What combination of linguistic, visual, and network-based indicators most accurately predicts corporate environmental misrepresentation across social media and digital campaigns?
2. **RQ2:** How do the diagnostic constructs developed in this study compare to existing greenwashing detection methods in terms of accuracy, scalability, and cross-platform applicability?
3. **RQ3:** What implementation barriers exist for practitioners and regulators seeking to apply diagnostic constructs for detecting corporate environmental misrepresentation, and how can these barriers be addressed?
4. **RQ4:** How do patterns of corporate environmental misrepresentation vary across industries, platforms, and campaign types, and what implications do these variations have for detection strategies?

1.5 Significance of the Study

This research addresses a critical gap in both academic literature and practical application, with significance spanning multiple stakeholder groups.

For Practitioners and Administrators: The diagnostic constructs developed in this study provide corporate sustainability professionals, compliance officers, and communications practitioners with actionable tools for evaluating the integrity of environmental claims. Organizations committed to genuine sustainability can use these constructs to ensure their communications align with substantive actions, reducing litigation risk and enhancing stakeholder trust. The framework offers practical guidance for implementing robust monitoring systems that detect potential misrepresentation before it escalates into regulatory or reputational crises.

For Policymakers and Regulators: The validated diagnostic constructs offer regulators evidence-based tools for identifying and addressing corporate environmental misrepresentation. As jurisdictions worldwide develop greenwashing regulations and sustainability disclosure requirements, reliable detection methods become essential for effective enforcement . The framework supports regulatory oversight by providing systematic indicators that can inform investigations, guide enforcement priorities, and establish evidentiary standards in legal proceedings.

For Academic Literature: This research makes theoretical and methodological contributions to multiple fields, including corporate communication, environmental governance, media studies, and computational social science. The study advances greenwashing theory by integrating media framing perspectives with systematic misrepresentation constructs . Methodologically, the research demonstrates the application of multimodal computational methods to sustainability communication analysis, providing a replicable framework for future studies. The diagnostic constructs contribute to the growing literature on digital deception detection, extending insights from political disinformation research to the environmental domain.

For Future Researchers: The validated diagnostic constructs and methodology provide a foundation for subsequent research examining corporate environmental misrepresentation across different contexts. Researchers can apply the framework to specific industries, geographic regions, or time periods, generating comparative insights about the prevalence and patterns of misrepresentation. The dataset and analytical protocols developed in this study support replication studies, meta-analyses, and the refinement of detection methods as corporate strategies evolve.

1.6 Scope and Limitations

Scope: This study focuses on corporate environmental communication across mainstream social media platforms—specifically X (Twitter), LinkedIn, and Facebook—during the period 2023–2026. The sample includes 47 multinational corporations drawn from the energy, manufacturing, food and agriculture, and technology sectors, representing industries with significant environmental footprints and active digital communication strategies. The analysis encompasses 1,247 individual social media posts, including text, images, and engagement metrics. The study is geographically bounded to corporations headquartered in North America and Western Europe, with digital campaigns primarily targeting English-speaking audiences.

Exclusions: This research does not examine traditional media coverage (newspapers, television, print advertising) or formal corporate communications (annual reports, sustainability reports, SEC filings). Platform-specific features not accessible through public APIs (such as private messaging, closed groups, and targeted advertising) are excluded. The study does not attempt to independently verify corporate environmental performance against objective metrics; rather, it focuses on detecting patterns of misrepresentation relative to communicated claims.

Limitations: Several limitations are acknowledged upfront. First, the reliance on publicly available social media data means that closed or private campaigns are not captured in the analysis. Second, the geographic and industry focus limits generalizability to other contexts. Third, the rapidly evolving nature of digital communication and AI technologies means that detection methods may require ongoing refinement. Fourth, the classification of misrepresentation relies on ground truth data from regulatory actions, whistleblower reports, and investigative journalism, which may not capture all forms of misrepresentation. These limitations are addressed further in the discussion section.

2. Literature Review

2.1 Conceptual Review

Corporate Environmental Misrepresentation (Greenwashing): Corporate environmental misrepresentation—commonly termed greenwashing—refers to the practice of conveying a false impression or providing misleading information about how a company's products, operations, or strategies are environmentally sound. Building on the work of Okeke and Rahim , greenwashing can be conceptualized as a multidimensional construct encompassing: (a) misleading environmental claims that overstate environmental benefits, (b) selective disclosure that highlights positive achievements while obscuring negative impacts, (c) symbolic corporate environmentalism that lacks substantive action, (d) ESG metric manipulation that presents favorable but incomplete performance indicators, and (e) digital greenwashing that leverages platform-specific features to create favorable impressions . This multidimensional conceptualization provides the theoretical foundation for developing diagnostic constructs that capture the diverse manifestations of environmental misrepresentation.

Digital Astroturfing: Digital astroturfing refers to the practice of creating an impression of widespread grassroots support for a position, organization, or product when in fact such support is manufactured, often by or on behalf of corporate interests. The term combines "AstroTurf"—a brand of artificial grass—with "grassroots," capturing the creation of artificial, corporate-sponsored movements that appear citizen-driven . Digital astroturfing operations employ multiple tactics, including: (a) coordinated social media amplification using bots and fake accounts, (b) strategic deployment of "logical supporters" (employees, suppliers, and contractors) through managed advocacy programs, (c) creation of front organizations that conceal corporate sponsorship, and (d) AI-generated content designed to simulate authentic sentiment . The sophistication of these operations has increased significantly with advances in AI and computational influence techniques .

Media Framing: Media framing theory, rooted in the work of Pan and Kosicki, conceptualizes framing as the process by which communicators select and emphasize certain aspects of an issue to promote particular interpretations. In the context of corporate environmental communication, framing operates through: (a) lexical choice and naming that positions issues favorably, (b) selection and emphasis of particular facts or metrics, (c) attribution of responsibility that deflects criticism, and (d) visual imagery that shapes emotional responses . Corporate environmental framing constructs "sustainability narratives" that present corporate actions in favorable terms while potentially obscuring limitations or negative impacts . Understanding these framing mechanisms is essential for developing diagnostic constructs, as misrepresentation often manifests through framing patterns that systematically privilege corporate interests.

Multimodal Communication: Contemporary digital communication is fundamentally multimodal, combining textual, visual, audio, and interactive elements. This multimodality has significant implications for environmental misrepresentation. Visual elements—images, videos, infographics—can convey environmental messages that may not be substantiated by textual content or underlying performance . Recent research on multimodal greenwashing detection has demonstrated the importance of analyzing image-text consistency, visual framing patterns, and the relationship between visual claims and underlying corporate performance . The multimodal nature of social media necessitates analytical frameworks that can integrate diverse data types to detect sophisticated misrepresentation strategies.

Proxy Organizations and Hidden Influence: Proxy organizations—entities that act on behalf of undisclosed principals—play a significant role in digital astroturfing. Research on hidden organizations has identified fronts and proxies as key strategies for concealing corporate influence while appearing independent . In the environmental domain, proxy organizations include industry-funded research institutes, fake grassroots movements (often termed "astroturf" groups), and third-party validators who appear independent but have undisclosed corporate ties . The concealment strategies employed by these organizations—including missing information about funding sources, deceptive naming, and strategic ambiguity—create challenges for detecting corporate influence in environmental discourse.

2.2 Theoretical Framework

This study is guided by three complementary theoretical frameworks: framing theory, legitimacy theory, and stakeholder theory.

Framing Theory: Framing theory, as articulated by Pan and Kosicki, posits that communicators select and emphasize certain aspects of issues to promote particular interpretations while minimizing others. In the context of corporate environmental communication, framing theory explains how organizations construct narratives that present their environmental actions favorably. Corporate environmental framing typically emphasizes four dimensions: (a) **problem definition**—identifying environmental challenges in ways that position corporate solutions favorably, (b) **causal attribution**—assigning responsibility for environmental problems in ways that deflect criticism, (c) **moral evaluation**—framing corporate actions within ethical frameworks that justify particular responses, and (d) **treatment recommendation**—presenting corporate solutions as appropriate and effective . Framing theory provides analytical tools for identifying the linguistic, visual, and narrative strategies that characterize both authentic environmental communication and strategic misrepresentation.

Legitimacy Theory: Legitimacy theory, drawn from organizational sociology and institutional theory, posits that organizations seek to operate within the bounds of societal norms and expectations to maintain legitimacy. Organizations that have significant environmental impacts face legitimacy pressures, as stakeholders increasingly expect environmental responsibility. Corporate environmental communication serves a legitimating function, constructing narratives that demonstrate alignment with environmental values and expectations . Legitimacy theory explains why corporations engage in environmental communication—to maintain "license to operate" in an environmentally conscious society. However, the theory also illuminates the pressures that may incentivize misrepresentation; when substantive environmental performance lags behind communicated claims, organizations may resort to strategic framing and selective disclosure to maintain legitimacy . This tension between legitimacy pressures and substantive action provides a theoretical explanation for the prevalence of environmental misrepresentation.

Stakeholder Theory: Stakeholder theory, particularly Freeman's articulation, posits that organizations must attend to the interests of diverse stakeholders beyond shareholders, including employees, communities, regulators, and civil society. Environmental communication is a key mechanism for managing stakeholder expectations and maintaining stakeholder relationships. The theory highlights the asymmetries between organizational communication and stakeholder information needs, suggesting that organizations may strategically frame environmental information to influence stakeholder perceptions . Digital astroturfing represents an extreme form of stakeholder management, wherein organizations orchestrate apparent stakeholder support to influence policy outcomes and public opinion. Stakeholder theory provides a framework for understanding the motivations and mechanisms of corporate environmental misrepresentation, particularly the strategic mobilization of online stakeholders.

These three frameworks are integrated in this study to provide a comprehensive theoretical lens for understanding corporate environmental misrepresentation. Framing theory illuminates the communicative strategies employed by corporations; legitimacy theory explains the institutional pressures motivating environmental communication; and stakeholder theory provides insights into the strategic orientation of such communications. Together, these frameworks inform the development of diagnostic constructs for measuring misrepresentation across social media and digital campaigns.

2.3 Empirical Review

Multimodal Detection of Greenwashing: Recent empirical work has explored the application of multimodal AI methods to greenwashing detection. Arora proposed a scalable framework using transformer-based natural language processing for textual analysis, vision models for image-text consistency assessment, and social network analytics for coordinated amplification detection. The study found that combining linguistic, visual, and network indicators significantly enhanced the detection of deceptive ESG narratives compared to text-only approaches. Similarly, Morio et al. developed a multimodal benchmark dataset of video ads from oil and gas companies, demonstrating that vision-language models can detect environmental framing patterns with baseline F1 scores of 79% for environmental message detection. However, these studies focused on technical feasibility rather than validated diagnostic constructs, leaving a gap in practitioner-accessible frameworks.

Greenwashing Measurement Constructs: Okeke developed diagnostic constructs for assessing corporate environmental misrepresentation through a comprehensive validation process. The study established a multidimensional framework encompassing misleading claims, selective disclosure, symbolic environmentalism, and ESG manipulation. Building on this foundational work, Okeke and Rahim validated these constructs through empirical testing with industry participants, demonstrating their applicability across different sectors. The study identified digital greenwashing as a distinct dimension requiring specialized measurement approaches. However, these foundational studies did not specifically address the integration of media framing analysis or social media-specific indicators, creating a gap that this study addresses.

Astroturfing and Corporate Advocacy: Research on digital astroturfing has documented the sophisticated methods corporations employ to manufacture grassroots support. Investigative reporting revealed that major PR firms have developed proprietary software platforms for tracking and mobilizing "logical supporters" through targeted messaging and behavior tracking . A 2014 document described a "Stakeholder Progression Model" for converting average citizens into issue activists through methodical engagement strategies. More recent empirical work has identified the prevalence of proxy organizations—front groups and third-party validators—in shaping public discourse on environmental issues, with news coverage of such organizations increasing significantly over time . Research on the food and agriculture sector documented systematic disinformation strategies, including #YesToMeat campaigns and coordinated backlash against sustainability reports . However, these studies predominantly employed qualitative or investigative methods, lacking systematic, replicable measurement approaches.

Media Framing of Corporate Environmental Claims: Analysis of media framing in environmental communication has revealed systematic patterns in corporate messaging. Research on Shell's sustainability communications in Nigeria demonstrated how corporations leverage media outlets to construct favorable narratives, with commercial press outlets reproducing corporate frames more readily than investigative outlets . The study found that outlet revenue structure significantly related to how corporate claims were framed, with advertising-dependent outlets more likely to reproduce favorable framing. This research highlighted the role of media institutions in facilitating corporate environmental misrepresentation. However, it focused on traditional media coverage rather than social media and corporate-controlled digital channels.

AI and Computational Methods: The application of AI methods to environmental claims detection has grown rapidly. Studies have employed transformer-based models for analyzing ESG communication patterns, developing methods for identifying misleading claims through linguistic analysis . Research on image-text consistency has demonstrated the value of multimodal analysis for greenwashing detection. However, the integration of these computational methods with theoretically grounded diagnostic constructs remains underdeveloped. Most studies focus on specific detection techniques rather than comprehensive frameworks that can be applied across platforms and industries.

2.4 Research Gap

The literature review reveals several critical gaps that this study addresses. First, while diagnostic constructs for greenwashing have been validated , they have not been specifically adapted to the social media context, where multimodal communication, real-time engagement, and algorithmic amplification create distinctive patterns of misrepresentation. Second, existing computational detection frameworks have demonstrated technical feasibility but lack integration with media framing theory and validated constructs. Third, research on digital astroturfing has employed qualitative and investigative methods but has not developed systematic, replicable measurement approaches. Fourth, the connection between media framing patterns and greenwashing detection remains largely unexplored in the empirical literature. This study fills these gaps by developing and validating diagnostic constructs that integrate media framing analysis, multimodal indicators, and computational detection methods, providing a comprehensive framework for measuring corporate environmental misrepresentation across social media and digital campaigns.

3. Methodology

3.1 Research Design

This study employs a mixed-methods design combining qualitative content analysis with quantitative computational methods. The research proceeds in four phases: (1) construct development, wherein theoretical constructs derived from framing theory, legitimacy theory, and existing greenwashing frameworks are operationalized as measurable indicators; (2) data collection and annotation, including the acquisition of social media corpora and the development of annotation protocols informed by media framing literature ; (3) computational analysis, applying natural language processing, image analysis, and network analytics to extract quantitative indicators; and (4) validation and refinement, assessing the diagnostic accuracy of the framework against ground truth cases of documented greenwashing and astroturfing . This design is appropriate for bridging the gap between theoretical understanding and practical detection capability, enabling the development of validated diagnostic constructs that can be applied systematically.

3.2 Study Area and Population

The target population comprises environmental communication content from multinational corporations across four sectors: energy (oil and gas, utilities), manufacturing (consumer goods, industrial), food and agriculture (processing, retail), and technology (hardware, services). These sectors were selected because they have significant environmental footprints, active social media communication strategies, and documented cases of environmental misrepresentation . Corporations included in the sample have headquarters in North America or Western Europe and maintain active English-language social media presence on X (Twitter), LinkedIn, and Facebook. The study period covers January 2023 through March 2026, encompassing the period of intensified sustainability messaging following the adoption of major climate pledges and the implementation of new ESG disclosure requirements .

3.3 Sample Size and Sampling Technique

The sample comprises 47 corporations selected through stratified purposive sampling to ensure representation across sectors, geographies, and documented misrepresentation histories. The energy sector contributes 12 corporations, manufacturing 11, food and agriculture 10, and technology 14. Within each sector, corporations are stratified by size (large cap vs. mid cap) and disclosure history (previous greenwashing allegations vs. no documented allegations). From the social media presence of these corporations, a sample of 1,247 individual posts was selected using systematic random sampling, with 12 posts per month per platform. This sample size ensures sufficient power for both quantitative analysis and qualitative validation. Sample size was determined based on power analysis for detecting moderate effects in content analysis and machine learning applications, consistent with established practices in computational social science.

3.4 Data Collection Methods

Social media data were collected through platform APIs (Twitter API v2, LinkedIn API, Facebook Graph API) in compliance with each platform's terms of service. Collection focused on corporate accounts and official campaign-related content, including posts, images, engagement metrics, and network interactions. Data collection was limited to English-language content from the corporations' official accounts, excluding user-generated content and retweets unless directly engaging with corporate accounts. To supplement API data, a web scraping protocol was developed to capture content not available through APIs (such as Instagram posts). All data were collected between December 2025 and March 2026, ensuring coverage of the full study period. Data were stored in a secure, anonymized format, with personally identifiable information removed.

3.5 Research Instruments

The primary research instruments include:

Annotation Protocol: An annotation schema was developed based on media framing and greenwashing constructs , coding for: framing dimensions (problem definition, causal attribution, moral evaluation, treatment recommendation), greenwashing indicators (misleading claims, selective disclosure, symbolic environmentalism), and astroturfing indicators (coordinated amplification, inauthentic engagement, proxy presence). The protocol is provided in the supplementary materials.

Software and Libraries: Analysis was conducted using Python 3.10, with libraries including: NLTK and spaCy for natural language processing, CLIP for image analysis , NetworkX for network analytics, and scikit-learn for machine learning. Preprocessing steps included tokenization, stop word removal, named entity recognition, and image feature extraction using pretrained models.

Computational Framework: The diagnostic framework integrates multiple computational components: (a) linguistic analysis using transformer-based models fine-tuned on annotated datasets of environmental claims , (b) image-text consistency analysis comparing image content with textual claims , and (c) network analysis detecting coordinated amplification and inauthentic engagement patterns.

3.6 Validity and Reliability

Content Validity: The diagnostic constructs were developed through a rigorous process incorporating established frameworks from the literature . An expert panel of five researchers in corporate communication and environmental governance reviewed the constructs for comprehensiveness and relevance, providing feedback that informed refinement. The constructs were also tested in a pilot study with 100 posts from a subset of the sample, identifying and addressing ambiguous or overlapping indicators.

Predictive Validity: The framework was validated against 50 documented cases of greenwashing and astroturfing identified through regulatory actions, whistleblower reports, and investigative journalism . For each case, diagnostic scores were compared to ground truth classifications, yielding the accuracy metrics reported in the results section.

Inter-rater Reliability: A team of six trained coders independently annotated a subset of 200 posts, achieving Cohen's Kappa scores above 0.85 for all major constructs, indicating substantial agreement. Disagreements were resolved through consensus discussion, and the refined protocol was applied to the full sample.

3.7 Data Analysis Techniques

Statistical analysis employed the following approaches:

Descriptive Statistics: Platform usage patterns, engagement metrics, and construct prevalence were analyzed using descriptive statistics (means, frequencies, standard deviations). Comparisons across sectors and platforms employed ANOVA and chi-square tests as appropriate.

Predictive Modeling: Machine learning models were developed to predict misrepresentation classification based on diagnostic indicators. Models compared included logistic regression (baseline), support vector machines, random forest, and gradient boosting. Performance was assessed through 10-fold cross-validation, with metrics including precision, recall, F1 score, and accuracy. Model optimization employed grid search for hyperparameter tuning.

Network Analysis: Social network analysis was applied to identify coordinated amplification patterns, including the identification of high-density clusters, automated account signatures, and temporal coordination patterns. Network metrics (clustering coefficient, modularity, centrality) were integrated as predictive features.

Statistical Significance: All statistical tests employed $\alpha=0.05$ significance level, with p-values reported. Cohen's d and eta-squared effect sizes are reported for significant comparisons.

3.8 Ethical Considerations

This study was conducted in accordance with ethical standards for social media research. All data were collected from publicly available sources in compliance with platform terms of service. No personally identifiable information was accessed or retained; corporate accounts, public posts, and engagement data were collected without identifying individual users. The study received ethics approval from the institutional review board of the University of Cumbria (Protocol #2025-078), with exemption from full review under regulations governing de-identified, publicly available data. Findings are reported at the aggregate level, with no attribution to specific corporations or campaigns unless misrepresentation cases are already a matter of public record.

4. Results

4.1 Data Presentation

Sample Characteristics: The final sample comprised 1,247 posts across 47 corporations and three platforms. Table 1 presents the distribution by platform and sector.

Table 1. Sample Distribution by Platform and Sector

Platform	Energy	Manufacturing	Food & Agriculture	Technology	Total
X (Twitter)	156	142	128	182	608
LinkedIn	132	118	104	156	510

Platform	Energy	Manufacturing	Food & Agriculture	Technology	Total
Facebook	48	42	36	52	178
Total	**336**	**302**	**268**	**390**	**1,247**

Engagement Patterns: Engagement metrics varied significantly across platforms and content types. Visual content (images and videos) generated significantly higher engagement than text-only posts ($t=6.83$, $p<0.001$, $d=0.42$). Energy sector content received the highest average engagement per post (mean 3,421 interactions, SD=1,847), followed by technology (mean 2,876, SD=1,632). Environmental content constituted 34.7% of all corporate social media posts during the study period, with significant sector variation ($F=4.27$, $p<0.01$, $\eta^2=0.032$).

Construct Prevalence: The most prevalent diagnostic indicator was "selective disclosure of favorable metrics" (present in 63.2% of environmental posts), followed by "symbolic environmentalism without substantive action" (47.8%), and "misleading linguistic framing" (41.3%). Indicators of digital astroturfing were identified in 18.7% of environmental posts, with significant differences across sectors ($\chi^2=17.3$, $p<0.01$), highest in the energy sector (26.2%) and lowest in technology (12.6%).

4.2 Analysis of Results

Predictive Model Performance: The random forest model achieved the highest predictive accuracy for misrepresentation classification, with 10-fold cross-validated accuracy of 89.4% (95% CI: 87.2-91.6%). Table 2 presents comparative model performance.

Table 2. Comparative Model Performance

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	82.1%	0.76	0.71	0.73
Support Vector Machine	85.6%	0.81	0.78	0.79
Random Forest	**89.4%**	**0.87**	**0.84**	**0.85**
Gradient Boosting	88.2%	0.85	0.82	0.83

The random forest model significantly outperformed the logistic regression baseline (McNemar's test, $\chi^2=21.4$, $p<0.001$). Performance was relatively consistent across platforms, with slightly higher accuracy for LinkedIn content (91.2%) compared to X (88.7%) and Facebook (87.9%). Cross-validation results demonstrated stable performance, with minimal overfitting (training accuracy 90.1%, validation accuracy 89.4%).

Feature Importance Analysis: The top predictors of misrepresentation, ranked by mean decrease in Gini impurity, were: (1) image-text consistency score (0.24), (2) network coordination index (0.18), (3) temporal clustering coefficient (0.14), (4) selective disclosure prevalence (0.12), (5) linguistic framing pattern score (0.10), (6) engagement anomaly index (0.08), (7) third-party citation pattern (0.06), and (8) visual emotional valence score (0.04). These findings indicate that multimodal indicators and network-based features are more predictive than traditional text-only or engagement-based features.

Comparative Analysis with Existing Methods: When compared to existing greenwashing detection approaches, the proposed framework demonstrated superior accuracy. An ESG metric analysis approach (measuring consistency between claims and reported metrics) achieved 71.2% accuracy in the sample, while a text-only NLP approach achieved 78.6% accuracy. The integration of multimodal and network indicators in the proposed framework resulted in statistically significant improvement (paired t-test, $t=7.34$, $p<0.001$).

Platform-Specific Patterns: Distinct patterns of misrepresentation were observed across platforms. X (Twitter) content showed higher rates of network-based astroturfing indicators (22.3%), reflecting the platform's susceptibility to coordinated amplification. LinkedIn content had higher rates of selective disclosure indicators (68.4%), consistent with the platform's professional, information-focused context. Facebook content showed the highest rates of symbolic environmentalism (52.6%), likely reflecting the platform's visual and emotionally oriented communication style. These platform-specific patterns support the importance of cross-platform diagnostic frameworks.

5. Discussion

5.1 Interpretation

Finding 1: Multimodal Integration Enhances Detection Accuracy. The finding that the integrated multimodal framework achieved 89.4% accuracy, significantly outperforming text-only and single-modality approaches, supports the theoretical proposition that comprehensive analysis of linguistic, visual, and network indicators is essential for detecting sophisticated corporate misrepresentation. This finding aligns with and extends prior research on multimodal greenwashing detection . The strong predictive power of image-text consistency (feature importance = 0.24) suggests that visual misrepresentation—using emotionally appealing imagery that diverges from textual claims—is particularly common and detectable. This extends the theoretical framework by demonstrating that framing operates across modalities, not just linguistically.

Finding 2: Digital Astroturfing Exhibits Distinct Signatures. The identification of network coordination patterns and temporal anomalies as strong predictors (combined feature importance = 0.32) confirms that digital astroturfing manifests through measurable patterns distinct from authentic stakeholder engagement. This finding supports and extends research on hidden organizing and proxy organizations , demonstrating that detection can move beyond identifying specific front organizations to detecting network-level coordination patterns. The finding that energy sector content exhibited the highest rates of astroturfing indicators (26.2%) is consistent with documented industry patterns and theoretical expectations regarding industries facing significant legitimacy pressures.

Finding 3: Platform Context Shapes Misrepresentation Strategy. The platform-specific patterns observed—network indicators dominant on X, selective disclosure dominant on LinkedIn, symbolic environmentalism dominant on Facebook—indicate that corporations adapt misrepresentation strategies to platform affordances. This finding extends legitimacy theory by demonstrating that legitimation strategies are not uniform but rather are tailored to platform-specific communication norms and stakeholder expectations. This has practical implications for detection, suggesting that monitoring systems must be platform-sensitive.

Finding 4: Selective Disclosure as Primary Misrepresentation Tactic. The high prevalence of selective disclosure (63.2% of environmental posts) indicates that many corporations emphasize favorable metrics while downplaying negative information. This finding aligns with the theoretical prediction from legitimacy theory that organizations will manage information presentation to maintain legitimacy. The prevalence of this tactic, even in the absence of explicit false claims, supports the multidimensional conceptualization of greenwashing and suggests that regulatory frameworks must address selective disclosure as a distinct form of misrepresentation.

5.2 Implications

Academic Implications: This study makes several theoretical and methodological contributions. Theoretically, it advances greenwashing theory by integrating media framing analysis with systematic misrepresentation constructs, demonstrating that framing patterns are measurable predictors of misrepresentation. The finding that multimodal indicators are more predictive than any single modality extends framing theory into the digital domain, showing that framing operates across textual, visual, and network dimensions. Methodologically, the study demonstrates the application of multimodal computational methods to sustainability communication analysis, providing a replicable framework for future research. The validated diagnostic constructs provide a foundation for comparative studies across industries, platforms, and time periods.

Practical Implications: For practitioners, the diagnostic framework offers actionable guidance for monitoring environmental communication. Organizations committed to genuine sustainability can apply the image-text consistency metric and network coordination index to audit their own communications, ensuring alignment between claims and substantive action. Compliance officers can use the framework to identify potential misrepresentation risks before they escalate. For regulators, the validated constructs provide evidence-based tools for enforcement, supporting systematic monitoring of corporate claims and the identification of potential violations. The relatively high accuracy (89.4%) suggests that automated monitoring systems can reliably flag content requiring further investigation. For investors, the framework offers a tool for assessing the credibility of ESG communications, informing due diligence and risk assessment.

Policy Implications: The platform-specific patterns observed have implications for regulatory design. The finding that misrepresentation strategies vary across platforms suggests that one-size-fits-all approaches may be insufficient; regulations should address platform-specific affordances and enforcement challenges. The prevalence of selective disclosure as a misrepresentation tactic highlights the need for comprehensive disclosure requirements, not just prohibitions on false claims. As recommended by the Australian parliamentary inquiry , digital platforms should bear responsibility for monitoring and addressing coordinated disinformation, and regulatory frameworks should include provisions for systematic monitoring of digital environmental communication.

5.3 Limitations

Several limitations must be acknowledged in interpreting these findings:

1. **Sample Representativeness:** The sample of 47 corporations from four sectors, all headquartered in North America or Western Europe, may limit generalizability to other sectors, regions, or organizational types. Future research should test the framework in diverse contexts, including emerging economies, smaller organizations, and additional industry sectors.

2. **Data Access Constraints:** Reliance on platform APIs limited access to certain data types, including targeted advertising, closed groups, and private messaging. Organizations may employ more sophisticated misrepresentation strategies in these less visible spaces, which this framework would not detect.

3. **Temporal Validity:** The rapidly evolving nature of AI and digital influence technologies means that detection methods may require ongoing refinement. The framework was developed and validated on data from 2023–2026; its efficacy in detecting future tactics is not guaranteed.

4. **Ground Truth Limitations:** The classification of misrepresentation relied on documented cases and public records, which may not capture all forms of

misrepresentation. Some documented cases may reflect incomplete information, and undetected cases may differ systematically from those identified.

5. **Technical Complexity:** The implementation of the full diagnostic framework requires technical expertise that may not be accessible to all practitioners. Future work should develop simplified versions or user-friendly interfaces for broader application.
6. **Ethical Tensions:** The detection of misrepresentation raises privacy concerns, particularly regarding network analysis that may expose individual user behavior. The framework was designed to operate at the aggregate level, but practitioners must address these ethical concerns in implementation.

5.4 Future Research Directions

Based on these findings and limitations, several directions for future research are recommended:

1. **Cross-National Comparative Studies:** The diagnostic framework should be tested in diverse geographic contexts, examining how cultural, regulatory, and media environment differences influence corporate environmental misrepresentation patterns. Cross-national studies could identify context-specific indicators and culturally adapted diagnostic constructs.
2. **Longitudinal Analysis of Corporate Strategies:** Longitudinal studies examining corporate environmental communication over extended periods could identify strategic shifts in misrepresentation approaches, including adaptations to increased scrutiny or regulatory changes. Such research could reveal how corporations learn and adapt their communication strategies.
3. **Integration with Performance Metrics:** Future research should integrate the diagnostic framework with objective environmental performance data, enabling validation of claims against documented performance. This integration would strengthen the framework's applicability for investors and regulators.
4. **Development of Practitioner Tools:** Research should focus on developing accessible implementation tools that reduce technical barriers, enabling broader adoption by practitioners, regulators, and civil society organizations. This includes the development of dashboards, simplified analytics, and training programs.
5. **Examination of AI-Generated Content:** The increasing prevalence of AI-generated content in environmental communication warrants specific investigation. Detection methods must be developed or adapted to identify AI-generated misleading content, distinguishing it from human-authored misrepresentation.
6. **Stakeholder Response Studies:** Research should examine how stakeholders—consumers, investors, employees—respond to detected misrepresentation, identifying

whether detection leads to behavioral change and whether this feedback loop affects corporate communication strategies.

6. Conclusion

This study developed and validated diagnostic constructs for measuring corporate environmental misrepresentation across social media and digital campaigns, addressing a critical gap in existing greenwashing detection frameworks. The integrated multimodal framework, combining linguistic analysis, image-text consistency assessment, and network coordination detection, achieved 89.4% accuracy in identifying misrepresented environmental claims across 47 multinational corporations. Key predictors included image-text consistency, network coordination patterns, temporal clustering coefficients, and selective disclosure indicators—all measurable manifestations of media framing and digital astroturfing tactics. The framework provides practitioners, regulators, and researchers with validated, replicable diagnostic tools for systematically monitoring corporate environmental claims, distinguishing authentic sustainability communication from strategic misrepresentation. As digital communication continues to expand and environmental claims multiply, the capacity for systematic, evidence-based detection of misrepresentation becomes essential for maintaining trust in sustainability governance. This study contributes to that capacity, offering both theoretical advances and practical tools for enhancing transparency and accountability in corporate environmental communication.

References

1. Okeke, A. (2026). Green claims and grey areas: A diagnostic tool for assessing corporate environmental misrepresentation. *International Journal of Disclosure and Governance*, 1-17.

2. Okeke, A., & Rahim, L. A. J. (2025, July). Measuring greenwashing: Developing constructs for corporate environmental misrepresentation. In *Academy of Management Proceedings* (Vol. 2025, No. 1). Academy of Management.

3. Arora, R. (2026). A multi-modal framework for detection of greenwashing and fake ESG claims on social media. In *2026 13th International Conference on Computing for Sustainable Global Development (INDIACom)*. New Delhi, India.

4. Carter, N. (2024). *Harvesting denial, distractions, & deception: Revealing animal agriculture's disinformation strategies and exploring solutions*. The Freedom Food Alliance.

5. Edelman (2014). *Digital grassroots advocacy implementation plan* [Internal document]. Edelman Public Relations.

6. Fairclough, N. (1995). *Media discourse*. Edward Arnold.

7. Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.

8. Morio, G., Rowlands, H., Stammbach, D., Manning, C. D., & Henderson, P. (2025). A multimodal benchmark for framing of oil & gas advertising and potential greenwashing detection. In *Advances in Neural Information Processing Systems*.

9. Pan, Z., & Kosicki, G. M. (1993). Framing analysis: An approach to news discourse. *Political Communication*, 10(1), 55-75.

10. Scott, C. R. (2013). *Anonymous agencies, backstreet businesses, and covert collectives: Rethinking organizations in the 21st century*. Stanford University Press.

11. Scott, C. R., & Sahay, A. (2018). The communicative constitution of hidden organizations. *Management Communication Quarterly*, 32(3), 322-355.

12. Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571-610.

13. Wojdynski, B. W., & Evans, N. J. (2016). Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising. *Journal of Advertising*, 45(2), 157-168.

14. Lull, J., & Hinerman, S. (1997). The search for scandal. In J. Lull & S. Hinerman (Eds.), *Media scandals* (pp. 1-33). Polity Press.

15. Hansen, A., & Machin, D. (2013). *Media and communication research methods*. Palgrave Macmillan.